# Probability and Statistics

# Table of contents

**Luck. Coincidence. Randomness. Uncertainty. Risk. Doubt. Fortune. Chance.**

- This course is about uncertainty, measuring and quantifying uncertainty, and making decisions under uncertainty.
- By uncertainty we mean the condition when results, outcomes, the nearest and remote future are not completely determined; their development depends on a number of factors and just on a pure chance.
- Probability is the logic of uncertainty.
- Probability is extremely useful in a wide variety of fields, since it provides tools for understanding and explaining variation, separating signal from noise, and modeling complex phenomena.

Areas of application:

- Statistics: Probability is the foundation and language for statistics, enabling many powerful methods for using data to learn about the world.
- Physics: Einstein famously said "God does not play dice with the universe", but current understanding of quantum physics heavily involves probability at the most fundamental level of nature. Statistical mechanics is another major branch of physics that is built on probability.
- Biology: Genetics is deeply intertwined with probability, both in the inheritance of genes and in modeling random mutations.
- Computer science: Randomized algorithms make random choices while they are run, and in many important applications they are simpler and more efficient than any currently known deterministic alternatives. Probability also plays an essential role in studying the performance of algorithms, and in machine learning and artificial intelligence.

- Meteorology: Weather forecasts are (or should be) computed and expressed in terms of probability.
- Gambling: Many of the earliest investigations of probability were aimed at answering questions about gambling and games of chance.
- Finance: At the risk of redundancy with the previous example, it should be pointed out that probability is central in quantitative finance. Modeling stock prices over time and determining "fair" prices for financial instruments are based heavily on probability.
- Medicine: The development of randomized clinical trials, in which patients are randomly assigned to receive treatment or placebo, has transformed medical research in recent years.

You will learn how to:

- how to evaluate probabilities, or chances of different results (when the exact result is uncertain),
- how to select a suitable model for a phenomenon containing uncertainty and use it in subsequent decision making,
- how to make optimal decisions under uncertainty.

References:


- Michael Baron, Probability and Statistics for Computer Scientists, 2nd edition, Chapman and Hall, 2014.
- J. K. Blitzstein, J. Hwang, Introduction to Probability, Chapman and Hall, 2015.
- course slides

# Experiments and random events. Sample space

**Definition**

*In probability theory, a random experiment means a repeatable process that yields a result or an observation.*

Examples:

- tossing a coin
- rolling a die
- extracting a ball form a box

**Definition**

*An elementary result of an experiment is an outcome of that experiment.*

**Definition**

*A sample space, denoted by $\Omega$, is the set of all possible outcomes of an experiment.*

The elements of a sample space are the elementary events.

# Random event

The mathematical framework of probability theory is set theory.

**Definition**

*A random event is a subset of the sample space.*

Examples: Rolling a die

- $A_i$: face $i$ appears, $i = \overline{1,6}$ (all the elementary events are events);
- B: an odd number appears;
- C: a number less than 4 appears;
- D: a number larger than 3 appears.

Express these events as subsets of the sample space.

# Certain event. Impossible event.

There are two special events for every experiment: the certain event and the impossible event.

**Definition**

*The certain event (denoted by $\Omega$) is an event which happens with certitude at each repetition of an experiment.*

Example: when tossing a coin, the event (one of the two faces appears) is a certain event of the experiment.

**Definition**

*The impossible event (denoted by $\emptyset$) is an event which never happens in a random experiment.*

Example: when extracting a ball from a box which contains only white balls, the event (a red ball is extracted) is an impossible event.

# Contrary events

In the case of rolling a die, let's denote by $A$ : an even number appears, and $B$ : an odd number appears. We observe that if the event $A$ does not take place, then the event $B$ takes place, and the other way round.

## Definition

*The* **contrary** *of an event $A$ is an event $B$ satisfying the property that, at any repetition of the experiment, if the event $A$ occurs then $B$ does not occur, and if the event $B$ occurs then $A$ does not occur. The events $A$ and $B$ are also called* **mutually exclusive events**.

If $B$ is the contrary of $A$ then $A$ is the contrary of $B$.
We denote the contrary of an event $A$ by $\overline{A}$.

# Compatible events

**Definition**

*The events A and B are compatible if they can occur simultaneously.*

Example: when throwing a die, the event A: an even number appears and the event B: one of the numbers 2 or 6 appears, are compatible.

**Definition**

*The events A and C are incompatible if they cannot occur simultaneously (disjoint events).*

Example: when rolling a die, the events A: an even number appears and C: an odd number appears are incompatible.

# Event implied by another event

**Definition**

*We say that the event $A$* **implies** *the event $B$ (or the event $B$ is implied by the event $A$) if the occurrence of the event $A$ means that the event $B$ occurs as well.*

Example: when rolling a die, the event A: face 1 appears implies the event B: an odd number appears. A is a subset of B.

Any event implies the certain event.

# Operations with events

**Definition**

*The union $A \cup B$ of two events $A$ and $B$ is the event which takes place when at least one of the events $A$ or $B$ occur.*

**Definition**

*The intersection $A \cap B$ of two events $A$ and $B$ is the event which occurs when both events $A$ and $B$ take place at the same time.*

Example: For the experiment of rolling one die, let's consider the following events:

$$A = \{1, 2, 5\}, \ B = \{3, 4, 5\}.$$

At least one of the events A and B take place if we obtain one of the faces 1, 2, 3, 4 or 5.

$$A \cup B = \{1, 2, 3, 4, 5\}$$

Both events take place if we obtain the face 5.

$$A \cap B = \{5\}$$

# Operations with events

## Definition

*The complement of an event A, denote $\overline{A}$, is the event that occurs every time A does not occur. It is the contrary of event A.*

Example: when rolling a die, if event A: an even number appears, then $\overline{A}$ : an odd number appers.

## Definition

*The difference $A \setminus B$ of two events A and B is the event which takes place when A occurs but B does not occur.*

Example: when rolling a die, if $A = \{1, 3, 5\}$ and $B = \{2, 4, 5\}$, then

$$A \setminus B = \{1, 3\}.$$

The difference $A \setminus B$ consists of the outcomes A which are not in B.

## Definition

*Events A, B, C... are disjoint (incompatible or mutually exclusive) if*

$$A \cap B \cap C \cap \ldots = \emptyset.$$

Disjoint events cannot occur at the same time.
Example: when rolling a die, if we consider the (elementary) events $A_1 = \{1\}$, $A_2 = \{2\}$, ..., $A_6 = \{6\}$, then

$$A_1 \cap A_2 \cap \ldots \cap A_6 = \emptyset.$$

## Definition

*Events A, B, C... are exhaustive if*

$$A \cup B \cup C \cup \ldots = \Omega.$$

Exhaustive events "cover" all the sample space $\Omega$.
Any event A and its complement $\overline{A}$ are disjoint and exhaustive.
Example: $A_1 \cup A_2 \cup \ldots \cup A_6 = \Omega$.

# Frequency

Let us consider an event A associated to a given experiment. We repeat the same experiment $N$ times and we denote by $\alpha$ the number of occurrences of the event A.

The number $\alpha \in \{1, 2, \ldots, N\}$ is called the **absolute frequency** of the event A.

# Frequency

Let us consider an event A associated to a given experiment. We repeat the same experiment $N$ times and we denote by $\alpha$ the number of occurrences of the event A.
The number $\alpha \in \{1, 2, \ldots, N\}$ is called the **absolute frequency** of the event A.

## Definition

*The number $f_n(A) = \frac{\alpha}{n}$ is called relative frequency of the event $A$.*

# Frequency

Let us consider an event A associated to a given experiment. We repeat the same experiment $N$ times and we denote by $\alpha$ the number of occurrences of the event A.
The number $\alpha \in \{1, 2, \ldots, N\}$ is called the **absolute frequency** of the event A.

## Definition

*The number $f_n(A) = \frac{\alpha}{n}$ is called relative frequency of the event $A$.*

Properties of the relative frequency:

- $0 \leq f_N(A) \leq 1$, for any $N \in \mathbb{N}^*$;
- $f_N(\Omega) = 1$, where $\Omega$ is the certain event;

# Frequency

Let us consider an event A associated to a given experiment. We repeat the same experiment $N$ times and we denote by $\alpha$ the number of occurrences of the event A.
The number $\alpha \in \{1, 2, \ldots, N\}$ is called the **absolute frequency** of the event A.

## Definition

*The number $f_n(A) = \frac{\alpha}{n}$ is called relative frequency of the event $A$.*

Properties of the relative frequency:

- $0 \leq f_N(A) \leq 1$, for any $N \in \mathbb{N}^*$;
- $f_N(\Omega) = 1$, where $\Omega$ is the certain event;

**In a long run (large number of repetitions N), the probability of an event A can be estimated by its relative frequency.**

$$\lim_{N \to \infty} f_N(A) = P(A)$$

# Equally possible events

**Definition**

*Consider two events $A$ and $B$ associated to an experiment. If there is no reason to suppose that the occurrence of one of the two events is favored with respect to the other, then we say that the two events are equally possible.*

Examples: We consider the experiment of tossing a coin. If the coin is fair, equally possible events are: obtaining heads or obtaining tails.

When rolling a die, we can obtain any of the six faces. If the die is fair, the events (1), (2), (3), (4), (5), (6) are equally possible.
The events $A = \{1, 2\}$ and $B = \{3, 4\}$ are also equally possible.

# Probability of an event. Classical defionition

If the sample space $\Omega$ consists of equally likely outcomes $\omega_1, \omega_2, \ldots, \omega_n$, the probability of each outcome is equal to the inverse of the number of outcomes from the sample space:

$$P(\omega_k) = \frac{1}{n}.$$

# Probability of an event. Classical defionition

If the sample space $\Omega$ consists of equally likely outcomes $\omega_1, \omega_2, \ldots, \omega_n$, the probability of each outcome is equal to the inverse of the number of outcomes from the sample space:

$$P(\omega_k) = \frac{1}{n}.$$

The probability of any event E consisting of k outcomes is:

$$P(E) = \frac{k}{n} = \frac{\text{nr. of outcomes in E}}{\text{nr. of outcomes in } \Omega}.$$

The outcomes forming an event E are also *favorable*. Therefore:

$$P(E) = \frac{\text{number of favorable outcomes to E}}{\text{number of total outcomes}}.$$

Example: We consider the experiment of rolling a die and the associated sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$. As these events are equally possible, what is the chance of occurrence of each of them ? The event A: an even number appears. What is the chance of occurrence of event A?

Example: Consider the experiment of tossing two coins and the sample space $A = \{(H, H), (H, T), (T, H), (T, T)\}$. As these events are equally possible,what is the chance of occurrence of each of them? What is the probability of the event B: at least one head appears?

# Axiomatic definition of probability

## Definition

*We call probability on the sample space $\Omega$ a function $P$ which associates to every event $A \in \mathcal{P}(\Omega)$ a number $P(A)$, called probability of $A$, such that:*

1. $P(A) \geq 0,\ \forall A \in \mathcal{P}(\Omega)$;
2. $P(\Omega) = 1$;
3. *if* $A_1, A_2, \ldots$ *are disjoint events* ($A_i \cap A_j = \emptyset, \forall i \neq j$), *then* $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

# Axiomatic definition of probability

## Definition

*We call probability on the sample space $\Omega$ a function $P$ which associates to every event $A \in \mathcal{P}(\Omega)$ a number $P(A)$, called probability of $A$, such that:*

1. $P(A) \geq 0, \forall A \in \mathcal{P}(\Omega)$;
2. $P(\Omega) = 1$;
3. *if $A_1, A_2, \ldots$ are disjoint events ($A_i \cap A_j = \emptyset, \forall i \neq j$), then* $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

The function $P : \mathcal{P}(\Omega) \to R_+$ is called probability measure.
The sample space $\Omega$ together with the probability measure $P$ (the pair $(\Omega, P)$) is called probability space.

# Properties of the probability measure P

The function $P$ has the following properties:

- $P(\emptyset) = 0$;

# Properties of the probability measure P

The function $P$ has the following properties:

- $P(\emptyset) = 0$;
- $P(\overline{A}) = 1 - P(A)$;

# Properties of the probability measure P

The function $P$ has the following properties:

- $P(\emptyset) = 0$;
- $P(\overline{A}) = 1 - P(A)$;
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ for any $A, B \in \mathcal{P}(\Omega)$ (this formula can be generalized for n events);

# Properties of the probability measure P

The function $P$ has the following properties:

- $P(\emptyset) = 0$;
- $P(\overline{A}) = 1 - P(A)$;
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ for any $A, B \in \mathcal{P}(\Omega)$ (this formula can be generalized for n events);
- If $A \subset B$ then $P(A) \leq P(B)$;

# Properties of the probability measure P

The function $P$ has the following properties:

- $P(\emptyset) = 0$;
- $P(\overline{A}) = 1 - P(A)$;
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ for any $A, B \in \mathcal{P}(\Omega)$ (this formula can be generalized for n events);
- If $A \subset B$ then $P(A) \leq P(B)$;
- for any $A_1, A_2, \ldots, A_n \in \mathcal{P}(\Omega)$ we have $P(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$.

# Independent events

## Definition

*The events $A$ and $B$ from $\mathcal{P}(\Omega)$ are called independent if* $P(A \cap B) = P(A) \cdot P(B)$.

# Independent events

**Definition**

*The events $A$ and $B$ from $\mathcal{P}(\Omega)$ are called independent if* $P(A \cap B) = P(A) \cdot P(B)$.

The events $E_1, E_2, ..., E_n$ are called independent if the occurrence of one of these events does not affect the probabilities of the others.
If events $E_1, E_2, \ldots, E_n$ are independent then
$P(E_1 \cap E_2 \cap \ldots \cap E_n) = P(E_1) \cdot P(E_2) \cdot \ldots \cdot P(E_n)$.

# Independent events

## Definition

*The events A and B from $\mathcal{P}(\Omega)$ are called independent if* $P(A \cap B) = P(A) \cdot P(B)$.

The events $E_1, E_2, ..., E_n$ are called independent if the occurrence of one of these events does not affect the probabilities of the others.

If events $E_1, E_2, \ldots, E_n$ are independent then
$P(E_1 \cap E_2 \cap \ldots \cap E_n) = P(E_1) \cdot P(E_2) \cdot \ldots \cdot P(E_n)$.

If A and B are independent events, then the events A and $\overline{B}$, $\overline{A}$ and B, $\overline{A}$ and $\overline{B}$ are also independent.

# Independent events

## Definition

*The events A and B from $\mathcal{P}(\Omega)$ are called independent if* $P(A \cap B) = P(A) \cdot P(B)$.

The events $E_1, E_2, ..., E_n$ are called independent if the occurrence of one of these events does not affect the probabilities of the others.

If events $E_1, E_2, \ldots, E_n$ are independent then

$P(E_1 \cap E_2 \cap \ldots \cap E_n) = P(E_1) \cdot P(E_2) \cdot \ldots \cdot P(E_n)$.

If A and B are independent events, then the events A and $\overline{B}$, $\overline{A}$ and B, $\overline{A}$ and $\overline{B}$ are also independent.

Example: if we toss two coins, the events A $=$(obtain heads on the first coin) and B $=$(obtain tails on the second coin) are independent.

**Multiplication rule.** If an experiment consists of k components (sub-experiments) for which the number of possible outcomes are $n_1, n_2, \ldots, n_k$ then the total number of experimental outcomes (the size of the sample space) is equal to $n_1 \times n_2 \times \ldots \times n_k$.

**Multiplication rule.** If an experiment consists of k components (sub-experiments) for which the number of possible outcomes are $n_1, n_2, \ldots, n_k$ then the total number of experimental outcomes (the size of the sample space) is equal to $n_1 \times n_2 \times \ldots \times n_k$.

**Permutations**

A permutation is a possible selection of k distinguishable objects from a set of n objects ($n \geq k$);the order of the sampled objects is important.

**Multiplication rule.** If an experiment consists of k components (sub-experiments) for which the number of possible outcomes are $n_1, n_2, \ldots, n_k$ then the total number of experimental outcomes (the size of the sample space) is equal to $n_1 \times n_2 \times \ldots \times n_k$.

**Permutations**

A permutation is a possible selection of k distinguishable objects from a set of n objects ($n \geq k$);the order of the sampled objects is important.
If the selection is performed:

- **with replacement** The number of possible ways to select the k objects is $P(n, k) = n^k$.
- **without replacement** The number of possible ways to select the k objects is $P(n, k) = n(n-1)(n-2)\ldots(n-k+1) = \frac{n!}{(n-k)!}$

**Combinations**
A combination is a possible selection of k indistinguishable objects from n objects ($n \geq k$); the order of the sampled objects is not taken into account.

**Combinations**

A combination is a possible selection of k indistinguishable objects from n objects ($n \geq k$); the order of the sampled objects is not taken into account. If the selection is performed:

- **without replacement** then the number of possible ways to select the k objects is

$$C(n,k) = C_n^k = \frac{n!}{(n-k)!k!}.$$

- **with replacement** then the number of possible ways to select the k objects is

$$C_r(n,k) = C_{n+k-1}^k = \frac{(n+k-1)!}{k!(n-1)!}.$$

# Conditional probability

## Definition

*The probability of the event $A$ conditioned by the occurrence of the event $B$ is denoted by $P(A|B)$ or $P_B(A)$ and is defined by*

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

*if $P(B) \neq 0$.*

We can also say "probability of A, given B".

# Conditional probability

### Definition

*The probability of the event $A$ conditioned by the occurrence of the event $B$ is denoted by $P(A|B)$ or $P_B(A)$ and is defined by*

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

*if $P(B) \neq 0$.*

We can also say "probability of A, given B".

Example: Consider rolling two dice. Let a be the number which appears on the first die and b the number appearing on the second die. What is the probability that $a + b > 8$, knowing that $b = 3$?

# Properties of conditional probability

- if $A_1$, $A_2$ are incompatible then $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B)$;
- if A and B are independent events then $P(A|B) = P(A)$ and $P(B|A) = P(B)$;
- if $A_1, A_2, \ldots, A_n$ are events such that $P(A_1 \cap A_2 \cap \ldots \cap A_n) \neq 0$ (they can occur simultaneously, i.e. they are compatible), then

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|(A_1 \cap A_2)) \cdot \ldots$$

$$\cdot P(A_n|(A_1 \cap \ldots \cap A_{n-1}));$$

- if $A_1, A_2, \ldots, A_n$ are independent events then

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2) \cdot \ldots P(A_n).$$

# Law of total probability

## Definition

*If the events $A_1, A_2, \ldots, A_n$ form a partition of the sample space $\Omega$ and $X \in \mathcal{P}(\Omega)$, then*

$$P(X) = \sum_{i=1}^{n} P(A_i) \cdot P(X|A_i).$$

# Law of total probability

## Definition

*If the events $A_1, A_2, \dots, A_n$ form a partition of the sample space $\Omega$ and $X \in \mathcal{P}(\Omega)$, then*

$$P(X) = \sum_{i=1}^{n} P(A_i) \cdot P(X|A_i).$$

Example: Three boxes have the following structure of white and black balls: (2,3), (4,1), (4,6). The event $A_i$ consists of choosing the box i. It is know that $P(A_i) = \frac{1}{3}$, $i = \overline{1,3}$. A box is randomly chosen and a ball is extracted. What is the probability that the extracted ball is black?

# Bayes' formula

## Definition

*If the events $A_1, A_2, \ldots, A_n$ form a partition of the sample space $\Omega$ and are the cause of the occurrence of an event $X$, then*

$$P(A_k|X) = \frac{P(A_k) \cdot P(X|A_k)}{\sum_{i=1}^{n} P(A_i) \cdot P(X|A_i)}.$$

# Bayes' formula

## Definition

*If the events $A_1, A_2, \ldots, A_n$ form a partition of the sample space $\Omega$ and are the cause of the occurrence of an event $X$, then*

$$P(A_k|X) = \frac{P(A_k) \cdot P(X|A_k)}{\sum_{i=1}^{n} P(A_i) \cdot P(X|A_i)}.$$

The probabilities $P(A_i)$, $P(X|A_i)$, $i = \overline{1, n}$ are called prior probabilities and $P(A_i|X)$ are called posterior probabilities. The event $X$ is called evidence.

# Bayes' formula

## Definition

*If the events $A_1, A_2, \ldots, A_n$ form a partition of the sample space $\Omega$ and are the cause of the occurrence of an event $X$, then*

$$P(A_k|X) = \frac{P(A_k) \cdot P(X|A_k)}{\sum_{i=1}^{n} P(A_i) \cdot P(X|A_i)}.$$

The probabilities $P(A_i)$, $P(X|A_i)$, $i = \overline{1, n}$ are called prior probabilities and $P(A_i|X)$ are called posterior probabilities. The event $X$ is called evidence. Example: Consider two boxes. The first one contains 2 white balls and 3 black balls, and the second contains 7 white balls and 5 black balls. The event $A_i$ means choosing the ith box. It is known that $P(A_1) = 0.4$ and $P(A_2) = 0.6$. We randomly choose one box and a black ball. What is the probability that this black ball is chosen from the second box?

# Bayes' formula

## Definition

*If the events $A_1, A_2, \ldots, A_n$ form a partition of the sample space $\Omega$ and are the cause of the occurrence of an event $X$, then*

$$P(A_k|X) = \frac{P(A_k) \cdot P(X|A_k)}{\sum_{i=1}^{n} P(A_i) \cdot P(X|A_i)}.$$

The probabilities $P(A_i)$, $P(X|A_i)$, $i = \overline{1, n}$ are called prior probabilities and $P(A_i|X)$ are called posterior probabilities. The event $X$ is called evidence.
Example: Consider two boxes. The first one contains 2 white balls and 3 black balls, and the second contains 7 white balls and 5 black balls. The event $A_i$ means choosing the ith box. It is known that $P(A_1) = 0.4$ and $P(A_2) = 0.6$. We randomly choose one box and a black ball. What is the probability that this black ball is chosen from the second box?
Answer: X=(a black ball is extracted), $P(A_2|X) = \frac{25}{49}$.